PATTERNS IN CHICAGO TRAFFIC

BY

AKSHAY HEMANT KETKAR

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Industrial Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Professor Richard Sowers

ABSTRACT

This thesis analyzes the taxi data released by the City of Chicago as announced on November 16, 2016 on [1]. The main objective of this thesis is to infer traffic trends in the city of Chicago through Time Series Analysis of taxi trips by observing the trends and general statistics of the hourly and daily variations in taxi trips in terms of mean trip counts and mean speeds. We also observe the hourly and daily taxi trip trends with respect to Chicago airports: O'Hare and Midway, in terms of the overall mean times to airports and mean times from major community areas, mean trip trends to and from airports, and pickup and drop-off statistics for major community areas. We also aim to study the impact of certain events on the traffic trends in the period of the dataset such as Christmas, Thanksgiving, New Year's Day and the Historic Snow Storm that occurred in the region from January 31, 2015 to February 2, 2015 [33]. In the end, we also introduce the Taxisim library [36], which contains algorithms for traffic estimation. This is introduced as a prototype method and is to be treated as a future assignment.

ACKNOWLEDGEMENTS

I take this opportunity to express my gratitude to my parents, Mr. Hemant Shriram Ketkar and Mrs. Medha Hemant Ketkar for their constant encouragement and support throughout my life without which, my journey to Master's would not have been possible. I am highly indebted to my adviser, Professor Richard Sowers for giving me the opportunity to work on a topic in my field of interest, for guiding me during the research and providing valuable inputs. I would also like to thank Professor Daniel Work for his inputs and regular support throughout the course of the project. Lastly, I am grateful to all the friends, peers and well-wishers who stood by me through thick and thin.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The city of Chicago released taxi GPS data for public on November 16, 2016 on [1]. Chicago residents and visitors took more than 27 million taxi rides in 2015, traveling 83 million miles and spending more than $400 million.

The Department of Business Affairs & Consumer Protection (BACP) of the City of Chicago assures the quality and safety of those rides. The BACP is also authorized to collect information on taxi rides, themselves. It does so through periodic reporting by two major payment processors believed to cover most taxis in Chicago. Based on these reports, they provided a dataset of over 100 million Chicago taxi rides, dating back to 2013. The timeline of the dataset is January 1, 2013 to November 1, 2016.

Each row in the dataset represents a unique taxi trip and consists of:

Trip ID: A Unique identifier for each trip

Taxi ID: Which Taxi provided the trip

Trip Start Timestamp: The time the trip started

Trip End Timestamp: The time the trip ended

Trip Seconds: Length of the trip in seconds

Trip miles: Length of the trip in miles

Pickup Census Tract: The census tract of the city of Chicago where the trip started

Drop-off Census Tract: The census tract of the city of Chicago where the trip started

Pickup Community Area: The community area of Chicago out of the 77 community areas where the trip started

Drop-off Community Area: The community area of Chicago out of the 77 community areas where the trip ended

Fare information: Such as trip fare, tips, tolls and extras, plus the total fare representing the total of all the components.

Payment type: Mode of payment such as Cash or Credit Card

Taxi company: The company providing the trip

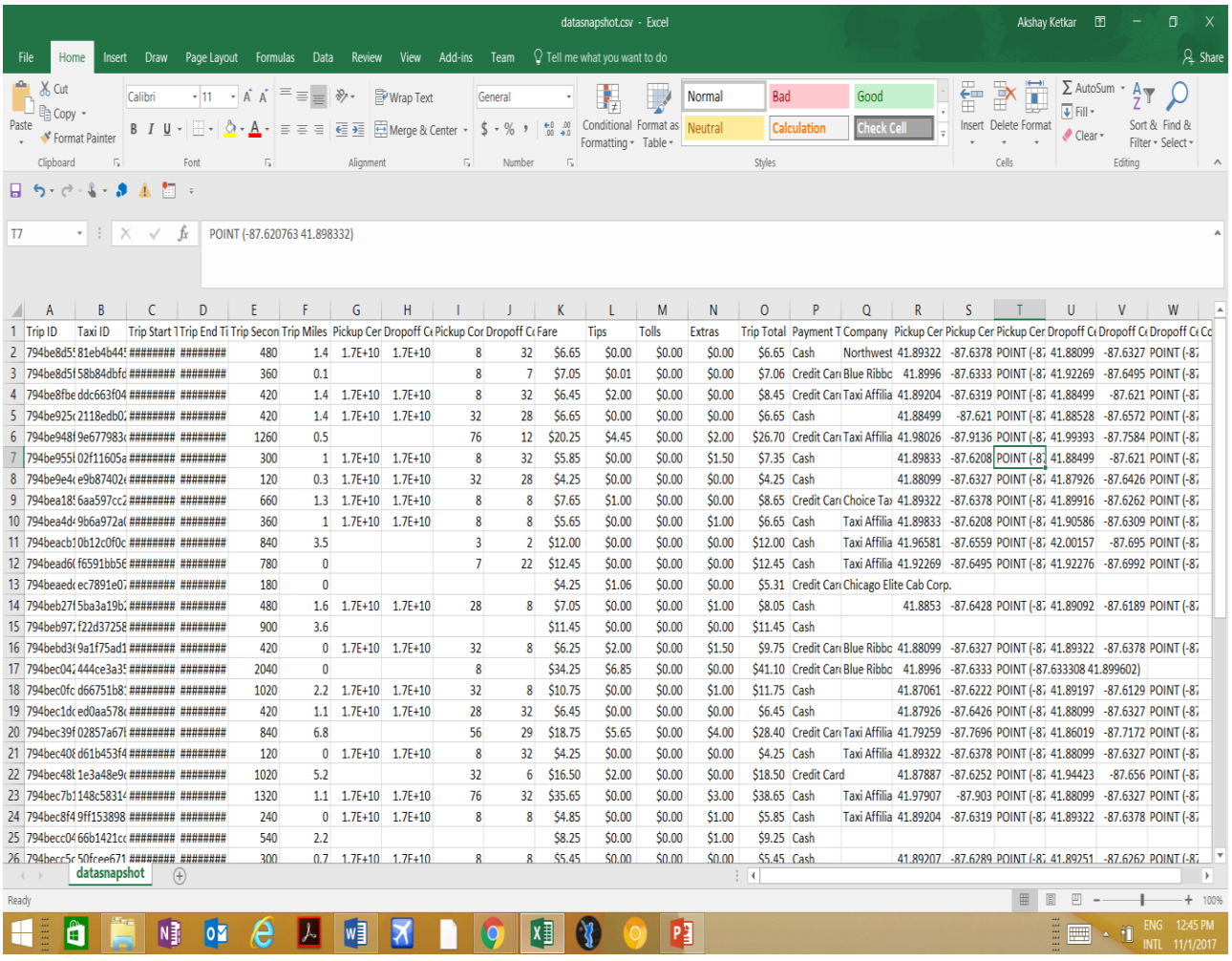


Figure 1: Snapshot of data

1.1 Privacy of Data:

Although taxi rides can be freely observed, it was realized that there could be privacy issues in publishing them. This issue was taken seriously and many measures were undertaken without unduly hampering the use of the data. The following measures were undertaken:

a) Delay: The taxi rides are not reported in real time, but many days or even months after taking place.

b) Masking of Medallion number: Taxi ID is not the real number on the taxi but an alias.

c) Masking of time: To address issues arising out of someone knowing that a trip took place at a certain time, the start and the end times have been rounded-up to the nearest 15 minutes.

d) Masking of location: As the publishing of exact time and exact location could affect privacy, the location has been approximated to the nearest Census Tract and Community Area Level.

1.2 Other Miscellaneous changes:

Some implausible values that could affect measures such as averages and data visualizations to an unreasonable degree have been removed. They are:

- Trip times less than zero or greater than 86,400 seconds (The number of seconds in a day) are removed.
- Trip lengths less than zero or greater than 3,500 miles are removed.
- If any component of the trip cost is less than $0 or greater than $10,000, all components of the trip cost are removed.

Explanation: A time of 86,400 seconds is one day. Even with breaks, it is unreasonable as it violates the number of hours a driver can drive per day. 3500 miles is the farthest in the United States that one could drive.

1.3 Chicago Community Areas:

As mentioned above, the City of Chicago is divided into 77 community areas. They are as follows:

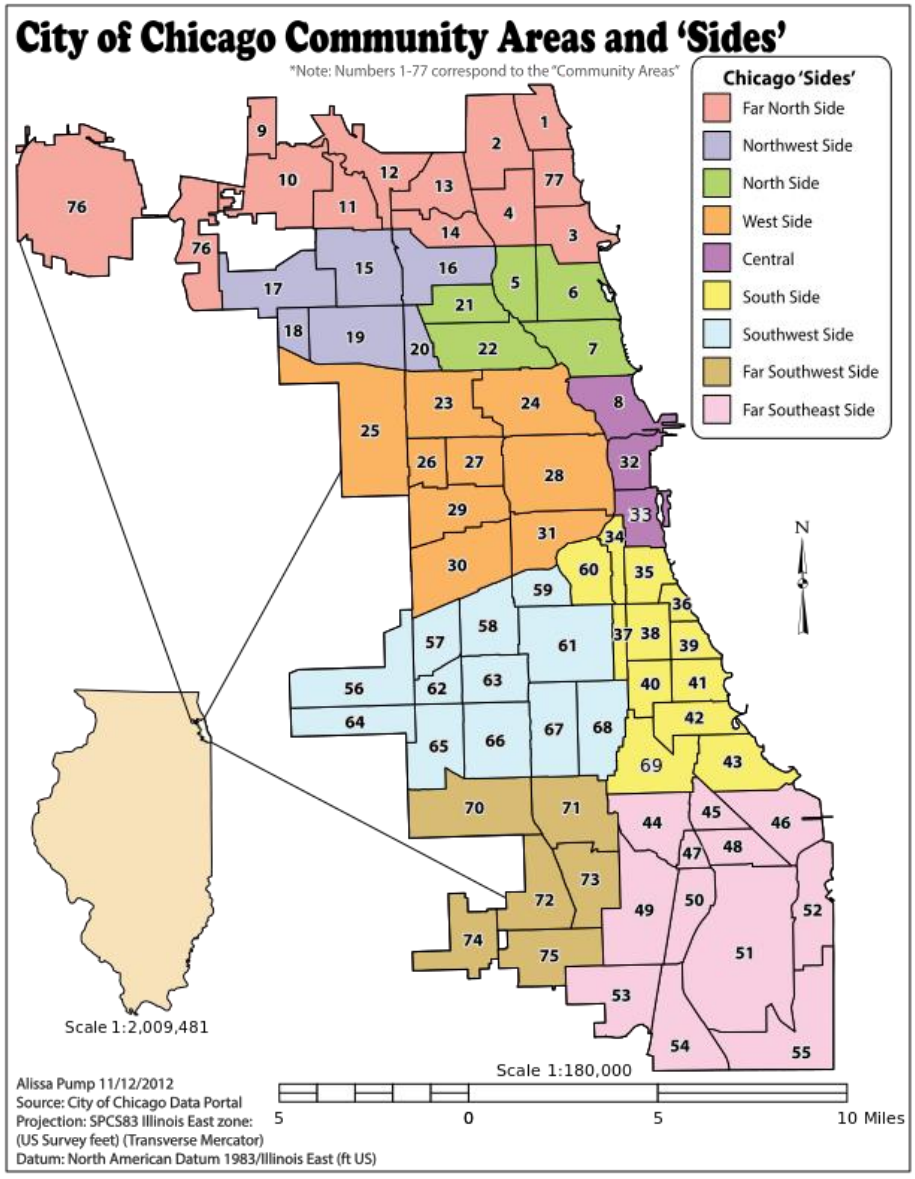


Figure 2[2]: Community area map of Chicago

1	ROGERS PARK	31	LOWER WEST SIDE	61	NEW CITY
2	WEST RIDGE	32	LOOP	62	WEST ELSDON
3	UPTOWN	33	NEAR SOUTH SIDE	63	GAGE PARK
4	LINCOLN SQUARE	34	ARMOUR SQUARE	64	CLEARING
5	NORTH CENTER	35	DOUGLAS	65	WEST LAWN
6	LAKE VIEW	36	OAKLAND	66	CHICAGO LAWN
7	LINCOLN PARK	37	FULLER PARK	67	WEST ENGLEWOOD
8	NEAR NORTH SIDE	38	GRAND BOULEVARD	68	ENGLEWOOD
9	EDISON PARK	39	KENWOOD	69	GREATER GRAND CROSSING
10	NORWOOD PARK	40	WASHINGTON PARK	70	ASHBURN
11	JEFFERSON PARK	41	HYDE PARK	71	AUBURN GRESHAM
12	FOREST GLEN	42	WOODLAWN	72	BEVERLY
13	NORTH PARK	43	SOUTH SHORE	73	WASHINGTON HEIGHTS
14	ALBANY PARK	44	CHATHAM	74	MOUNT GREENWOOD
15	PORTAGE PARK	45	AVALON PARK	75	MORGAN PARK
16	IRVING PARK	46	SOUTH CHICAGO	76	O'HARE
17	DUNNING	47	BURNSIDE	77	EDGEWATER
18	MONTCLARE	48	CALUMET HEIGHTS		
19	BELMONT CRAGIN	49	ROSELAND		
20	HERMOSA	50	PULLMAN		
21	AVONDALE	51	SOUTH DEERING		
22	LOGAN SQUARE	52	EAST SIDE		
23	HUMBOLDT PARK	53	WEST PULLMAN		
24	WEST TOWN	54	RIVERDALE		
25	AUSTIN	55	HEGEWISCH		
26	WEST GARFIELD PARK	56	GARFIELD RIDGE		
27	EAST GARFIELD PARK	57	ARCHER HEIGHTS		
28	NEAR WEST SIDE	58	BRIGHTON PARK		
29	NORTH LAWNDALE	59	MCKINLEY PARK		
30	SOUTH LAWNDALE	60	BRIDGEPORT		

Table 1[2]: List of Chicago community areas

2. LITERATURE REVIEW AND RELATED WORK

Brian Donovan in his thesis [3] proposed a method to quantitatively measure the resilience of transportation systems using GPS data from taxis. By computing the historical distribution of pace (normalized travel times) between various regions of a city and measuring the pace deviations during an unusual event and applying it to a dataset of nearly 700 million taxi trips in New York City, which is used to analyze the transportation infrastructure resilience to Hurricane Sandy.

Umang Patel in his paper [4] highlighted, the prevailing focus on the dataset of NYC taxi trips and fare. As during the early 2000s the taxi services exponentially increased and the data captured by NYC was in GBs and was very difficult to analyze manually, he developed a way to effortlessly analyze the thousands of GB within a fractions seconds and expedite the process. Using his method, the data could be analyzed for several purposes like avoiding traffics, lower rate where services are not functioning more frequency than a cab on crown location and many more. This information can be used by numerous authorities and industries such as government and Uber for their own purpose.

Omer et al. [5] proposed a method which measures the resilience of a road-based transportation network in terms of travel times between cities.

Chang et al. [6] evaluated a post-earthquake transportation network in terms of accessibility and coverage. This is partly based on an accessibility metric devised by Allen et al. [7], which considers travel times between various regions of a city. Thus, travel time is a standard quantity on which to measure resilience.

Another study measures temporal patterns in the density of taxi pickups and drop-offs to identify the social function of various city regions [8]. They point out that unusual output can be used to detect events like holidays.

Chen [9] specifically focuses on identifying anomalous taxi trajectories, to detect fraud or special events.

Ferreira et al. [10] created a graphical querying tool which can be used to count taxi trips between arbitrary geometrical regions as a function of time. They noted the drop in the frequency of taxi trips during Hurricane Sandy and Hurricane Irene, pointing out that the Irene related drop was more significant, but the Sandy-related drop was longer lasting.

Todd W. Schneider in [11], performed and an open-source exploration of the city's neighborhoods, nightlife, airport traffic, and more, through the lens of publicly available taxi and Uber data. He answered many questions such as How bad is the rush hour traffic from Midtown to JFK? Where does the Bridge and Tunnel crowd hang out on Saturday nights? What time do investment bankers get to work? How has Uber changed the landscape for taxis? And could Bruce Willis and Samuel L. Jackson have made it from 72nd and Broadway to Wall Street in less than 30 minutes?

3. METHODOLOGY

We try to follow the Knowledge Discovery in Databases or KDD methodology as described by Prof. Jiwaei Han [12]

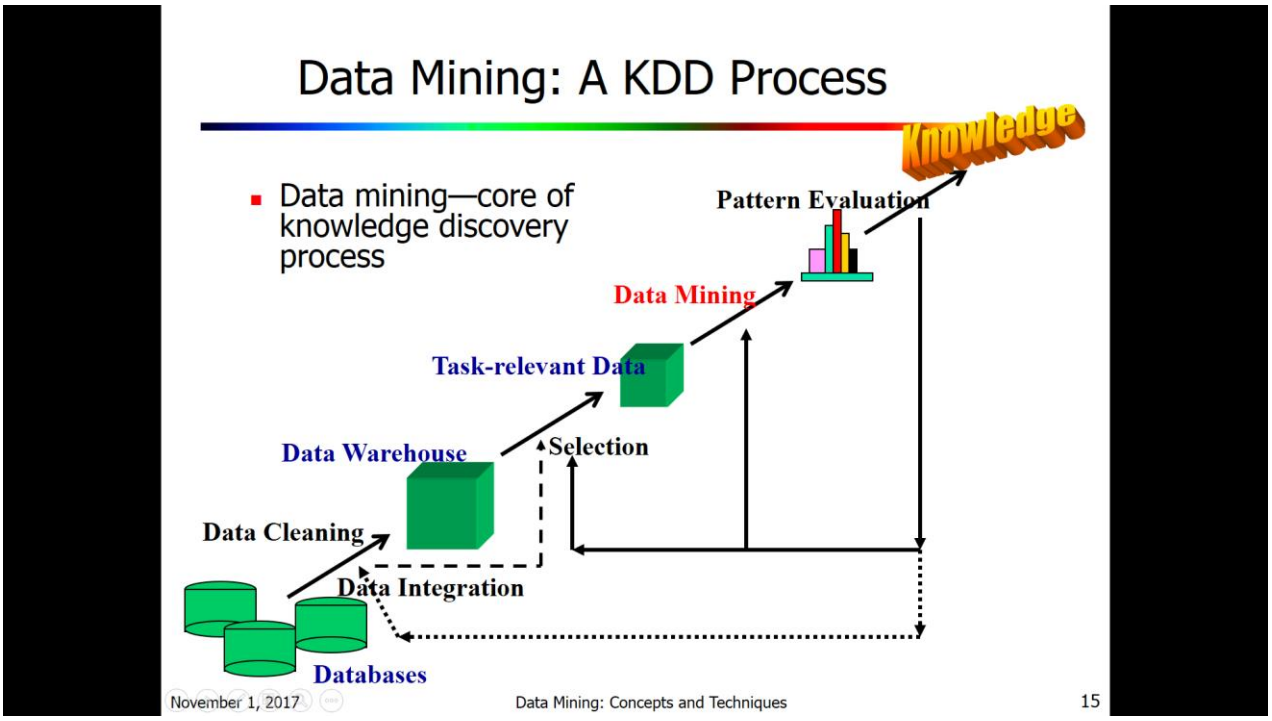


Figure 3[12]: KDD process

We draw inspiration from this process through the steps: Data Filtering, Fitting distribution to final data using Maximum Likelihood Estimation, Attribute Extraction, Data Integration and Selecting Task-Relevant Data for doing Data Mining and finding patterns.

3.1 Data Filtering:

Although, the dataset consists of about a 100 million unique taxi trips, it is important to extract only the data, which is complete and free from erroneous or missing entries. Also, it is important to detect outliers and set a threshold beyond which to filter out data so that there is no impact on certain measures such as average. It was decided to consider only those rows which passed the filtration test. This ensures consistency and availability of all information across attributes so that they can be grouped together for performing analysis.

3.1.1 Identification of critical attributes: For filtering the data, it was important to identify the attributes, which were critical from the point of view of the scope of the research. The following attributes were identified as being critical:

1) Trip Start time stamp

2) Trip End Time Stamp

3) Trip Seconds

4) Trip miles

5) Pickup Community Area

6) Drop-off Community Area

3.1.2 Not critical but required attributes: The other attributes being used for analysis are

1) Fare, tips, tolls, extras, total

2) Payment type

3) company.

4) Pickup and Drop-off Latitudes and Longitudes(Masked)

Rows with blank or missing entries in these attributes are not filtered out.

3.1.3 Initial Filtering: As recommended by Prof. Dan Work and Prof. Richard Sowers, it was decided to leave out the rows if the Start and the End Time Stamps were blank, if the Trip seconds and Trip miles were either blank or zero. If the Pickup and Drop-off Community Area columns were blank, they were left out. In addition, speeds in MPH were computed for each row and it was checked if the speed was greater than 70 MPH. The 70 MPH threshold was fixed considering a reasonable estimate of speed as most of the streets of the City of Chicago have a speed limit of 30 MPH. The initial filtering results in the elimination of about 40.31 % of the data and the size reduces to 61.26 million from 102.63 million.

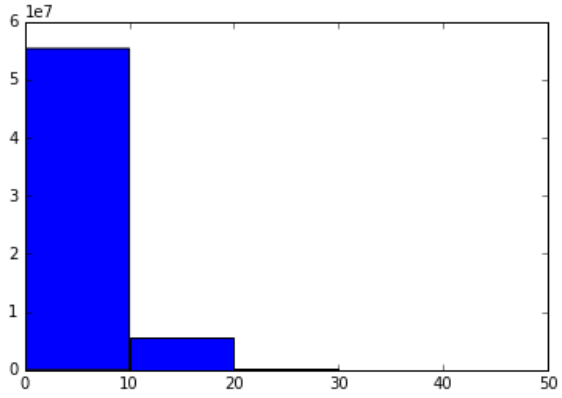


Figure 4a: Histogram of miles up to 50

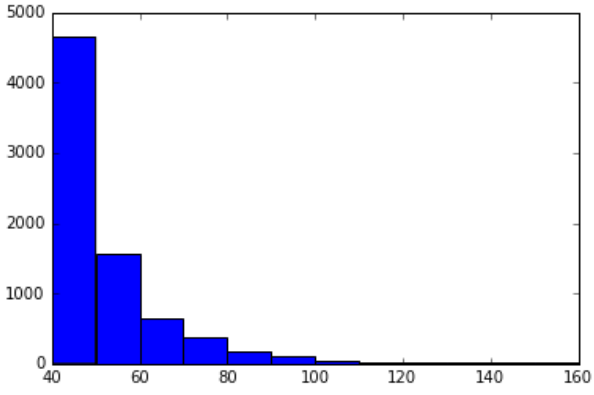


Figure 4b: Histogram of miles 40-160

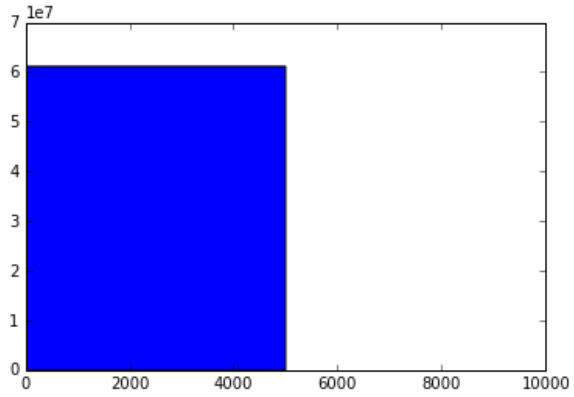


Figure 5a: Histogram of time up to 10000 seconds

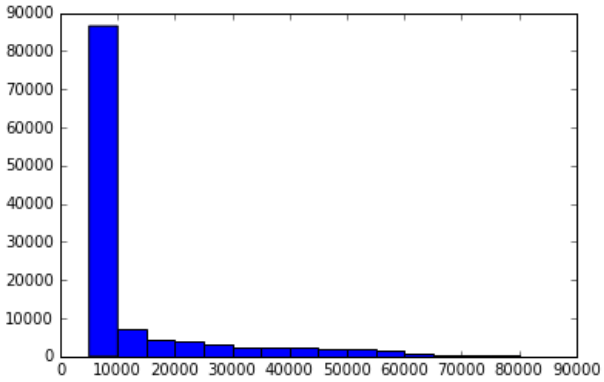
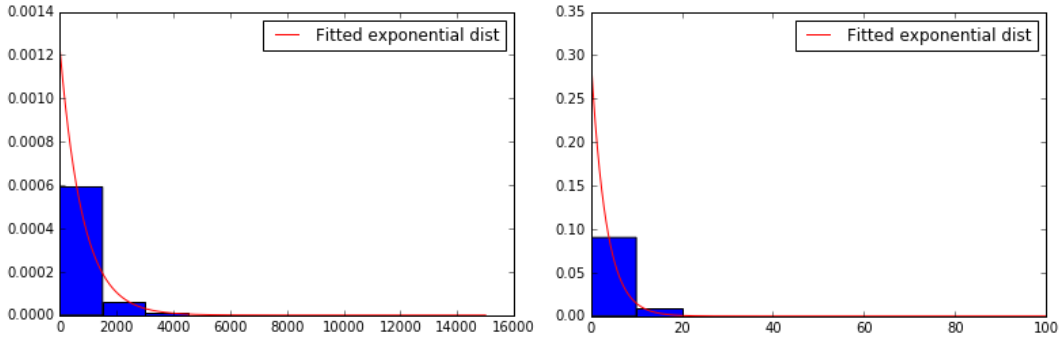


Figure 5b: Histogram of time 5000 and above

3.1.4 Final Filtering: Upon observing the histograms of trip miles and trip times, it is observed that for miles, the bulk of the values lie between 0 and 20 miles and for time, the bulk of the values lie between 0 and 5000 seconds. However, considering that there are considerable number of values even beyond 50 miles for distance and 10000 seconds for time, we fix the threshold for carrying out final filtering out miles at 100 miles and for time at 15000 seconds. This is reasonable considering that a miniscule number of trips could be unusually long both in terms of time and distance. This process results in a further reduction of data from 61.26 million to 61.23 million or 0.02%.

3.2 Fitting distribution to final data using Maximum Likelihood Estimation:



Figures 6: Pdfs of fitted exponential distributions for time(normalized) and miles(normalized) respectively

Maximum Likelihood Estimation: It is a method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters [13]. MLE aims to maximize the likelihood function:

$L = f(x_1, x_2, \ldots, x_n | \Theta) = f(x_1 | \Theta) \times f(x_2 | \Theta) \times f(x_n | \Theta)$

We work with the Natural Log of the likelihood function as it is easier to work with. We obtain the MLE for the parameter of the distribution by taking derivative of the log-likelihood function and setting it equal to zero.

Exponential distribution for the data: We now fit an exponential distribution [14] for the final data using Scipy [15] as the data closely resembles what an exponential distribution looks like. The MLE for an exponential distribution parameter (λ, shape parameter) is given by: $\lambda = n/\sum_{i=1}^{n} x_i$, where n is the size of the data and $\sum_{i=1}^{n} x_i$ is the sum of all values of the data. The MLE for λ for time is 0.0012 and miles is 0.3012

3.3 Attribute Extraction: Each of the six critical attributes mentioned above are extracted using the following steps:

Step 1: Import Taxi_trips.csv file

Step2: Read file

Step3: Skip header

Step4: Initialize attribute container

Step5: For every row in the file, apply filter conditions as above (Initial + Final). If row fails, continue, else assign appropriate attribute entry to attribute container.

Step7: Export attribute container to a file

Step8: Repeat Step1 to 7 for all required attributes

3.4 Data Integration:

Once the required attributes are extracted, the next step is to integrate the requisite attributes for analysis. For instance, if we require to find the taxi pick-up count on December 25, 2013 for community area 8, we need to integrate the attributes such as the Start time stamp and Pickup community area. Similarly, for observing the taxi trip count for trips between community area 8 and community area 76, we need to add one more attribute from the above: End Community Area. We could then export the integrated data to a file for future use.

3.5 Selecting Task-Relevant Data for doing Data Mining and finding patterns:

Often, data integration takes a lot of memory and it becomes necessary to export the integrated data to a file and restart the kernel. We, thereby load the integrated data from the file, so that we can perform certain operations such as slicing [16] and performing roll-up [16] operations on the data. For example, to find the total number of taxi trips from community area 32 by year, we must first slice the data for community area 32, then group the data by timestamp and then summarize the data for each year. We can then visualize the data after summarization to find interesting patterns.

3.6 Tools used:

We make extensive use of the high utility Pandas [17] package for Data Integration and Selecting Task-Relevant Data for doing Data Mining and finding patterns. We use the following pandas' commands for the above:

pandas.read_csv()[18]: To read csv file

pandas.to_datetime()[19]: Used to convert timestamps to pandas readable timestamps

pandas.Data_frame.sort_index()[20]: Used to sort according to the values of an index attribute. Can be used to sort data according to timestamps.

pandas.concat()[21]: Used to join different attributes together either columnwise(axis=1) or rowwise(axis=0)

pandas.DataFrame.groupby()[22]: Used to group by levels of an attribute. Can be used in conjunction with operations such as mean(), sum() or count. For example, to find the mean of trip time by Start Time Stamps, we use the following command:

DataFrame1= Dataframe_original.groupby(["Start Time Stamp"]) ["time"].mean()

4. RESULTS

4.1 Some important terms and definitions:

1) Stst: Start timestamp

2) Etst: End timestamp.

3) Hour of the day: 0(12 am) and 23(11 pm) [23]

4) Day of the week: 0(Monday), 6(Sunday) [24]

5) Mean trip counts (Includes Pickups and Drop-offs): Mean of the trip counts of all the recorded timestamps for a period. The lowest unit of timestamp is 15 minutes. For e.g. The overall mean trip count for an hour of the day will be the mean of counts for all recorded timestamps in that hour. Similar logic applies for a day.

6) Mean time: Mean of the mean time values in seconds for all the recorded timestamps for a period. The lowest unit of timestamp is 15 minutes. For e.g. The overall mean time in sec an hour of the day will be the mean of mean time values for all recorded timestamps in that hour. Similar logic applies for a day.

6) Mean speed: Mean of the mean speed values in MPH for all the recorded timestamps for a period. The lowest unit of timestamp is 15 minutes. For e.g. The overall mean speed in MPH for an hour of the day will be the mean of mean speed values for all recorded timestamps in that hour. Similar logic applies for a day.

7) Mean total fare: Mean of the total fare collected in $ for all the recorded timestamps for a period. The lowest unit of timestamp is 15 minutes. For e.g. The overall mean total fare collected in $ for an hour of the day will be the mean of the total fare values for all recorded timestamps in that hour. Similar logic applies for a day.

8) Pearson Correlation coefficient [25] and Cross-Correlation [26]: The Pearson product-moment correlation coefficient is a measure of the strength of the linear relationship between two variables. It is given by: $\rho_{X,Y} = COV(X,Y)/\sigma_X \sigma_Y$, where σ_X, σ_Y are std. dev of variables X and Y respectively. Cross-correlation is a measure of similarity of two series as a function of the displacement of one relative to the other, also known as sliding dot product. We use Pearson correlation coefficient instead of cross-correlation as we are comparing trends for similar time periods. Cross-correlation with a lag of 0 is equal to Pearson correlation.

9) Standardization: [27] Data standardization is the process of bringing data into a common scale and format that allows for collaborative research, large-scale analytics, and sharing of sophisticated tools and methodologies. The simplest method is rescaling the range of features to scale the range in [0, 1] or [−1, 1]. The formula is given by $x' = (x - \min(x))/(\max(x)-\min(x))$, where x is an original value, x' is the normalized value.

4.2 Summary Statistics:

Sr. No.	Area pairs	% of total	Description
1	(8, 8)	11.01	Within Near North Side
2	(8, 32)	7.50	Near North Side to The Loop
3	(32, 8)	7.46	Loop to the Near North Side
4	(32, 32)	4.65	Within Loop
5	(8, 28)	3.6	Near North Side to Near West Side
6	(32, 28)	3.00	Loop to Near West Side
7	(28, 8)	2.86	Near West Side to Near North Side
8	(8, 7)	2.45	Near North Side to Lincoln Park
9	(28, 32)	2.25	Near West Side to Loop
10	(8, 6)	2.05	Near North Side to Lake View
11	(76, 8)	1.76	O'Hare to Near North Side
12	(6, 6)	1.75	Within Lake View
13	(8, 24)	1.70	Near North Side to West Town
14	(7, 8)	1.66	Lincoln Park to Near North Side
15	(6, 8)	1.41	Lake View to Near North Side

Table 2: Frequency of trips between community areas

Sr. No.	Community area	% trips
1	Near North Side	29.38
2	The Loop	18.94
3	Near West Side	9.76
4	Lake View	7.96
5	Lincoln Park	7.13
6	West Town	5.11
7	O'Hare	3.78
8	Near South Side	3.21
9	Uptown	2.12
10	Logan Square	2.03

Table 3a: Average daily pickups

Sr. No.	Community area	% trips
1	Near North Side	33.78
2	The Loop	21.94
3	Near West Side	8.79
4	Lake View	7.60
5	Lincoln Park	6.13
6	O'Hare	5.29
7	West Town	3.94
8	Near South Side	2.43
9	Uptown	1.78
10	Garfield Ridge	1.65

Table 3b: Average daily drop-offs

Sr. No.	Company	%total
1	Taxi Affiliation Services	27.26
2	Dispatch Taxi Affiliation	11.25
3	Choice Taxi Association	6.07
4	Northwest Management LLC	3.49
5	Blue Ribbon Taxi Association Inc.	2.40
6	KOAM Taxi Association	2.12
7	Top Cab Affiliation	1.25
8	Chicago Medallion Leasing INC	0.37
9	Chicago Medallion Management	0.24

Table 4: Trip percentage by company

Mode	%total
Cash	62.52
Credit Card	36.84
No Charge	0.47
Unknown	0.09
Dispute	0.05
Pcard	0.02
Prcard	0.0091

Table 5: Mode of Payment

Frequent trips (Table 2): Out of 5274 combinations of Community Areas, the highest proportion of trips occur within the community area 8(Near North Side). This is not surprising considering the location of major landmarks and attractions such as the Navy Pier, John Hancock tower, the Trump tower. Also, many countries have their consulates located here. It is also home to many businesses and trade missions. Among the other frequent trips, trips within the Loop, between the Loop and Near North Side, between the Near North Side and Near West Side and between the Near West Side and the Loop are observed. It is not surprising because most of the areas are located near Downtown Chicago and are home to several popular tourist attractions, hotels and restaurants. Overall, the top 15 pairs as per Table 2 account for nearly 55% of the total trips.

Average daily pickups and drop-offs (Tables 3a, 3b): Community area 8 or Near North-Side, community area 32 or The Loop and community area 28 or Near West-Side top the chart for the highest percentage (More than 50) of average daily pickups and drop-offs. It is not surprising considering the proximity with Downtown Chicago. As mentioned above, most of the businesses, landmarks, tourist attractions, restaurants and other places of economic importance lie in these community areas.

Most preferred Taxi operators (Table 4): The most popular taxi operator has been Taxi Affiliation Services, followed by Dispatch Taxi Affiliation and Choice Taxi Association. Overall the top 9 operators account for nearly 54.45% of the trips considering we have information for only about 55% of trips (Rest are blank)

Mode of payment (Table 5): The most preferred mode of payment is Cash and Credit Card. Together they account for nearly 99.36% of the payments. Some trips (0.47%) incurred no charge. The rest are either Unknown, disputed or made by Pcard or Prcard.

Co-relation between base fare and tips: The Pearson correlation coefficient is 0.1124, which indicates a weak co-relation between base fare and tips. Hence, tips are entirely subjective and not a fixed percentage.

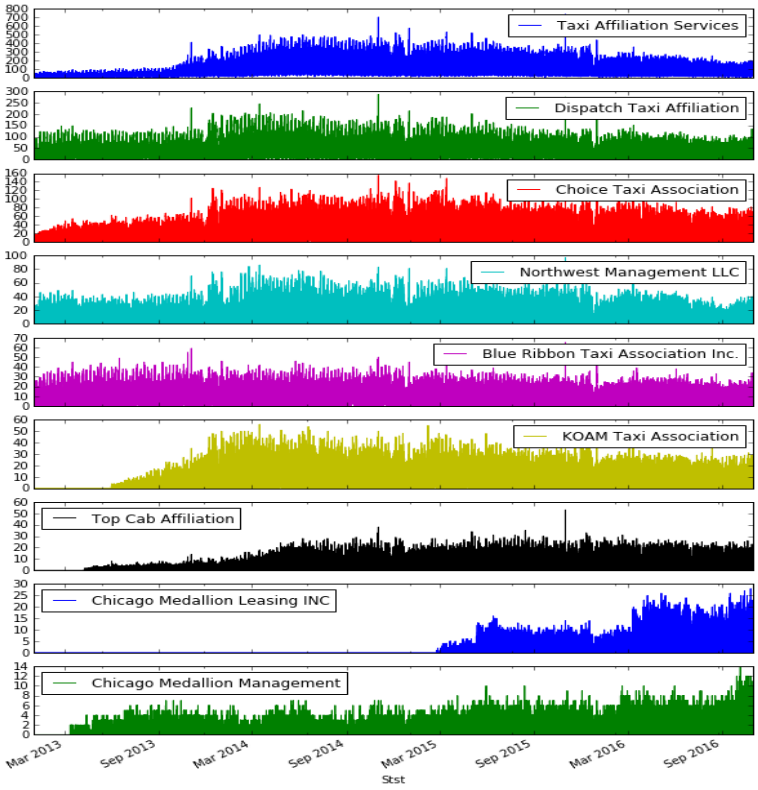


Figure 7: Year-wise taxi trip trends by companies

Taxi trip trends by companies: Out of the top 9 taxi operators, Taxi Affiliation Services, Dispatch Taxi Affiliation, Choice Taxi Association and Northwest Management LLC show a bulge between late 2013 and late 2015, followed by a dip. Blue Ribbon Taxi Association shows a somewhat constant trend followed by a gentle dip post late 2015. KOAM Taxi Association showed a sharp rise from late 2013, followed by a gentle dip towards 2016. Top Cab Affiliation rose post March 2014 and remained somewhat stable. Chicago Medallion Leasing INC showed a sharp rise post March 2016 and Chicago Medallion Management shows a gentle rise post March 2015, followed by a steep rise post September 2016.

4.3 Hourly and Daily trends:

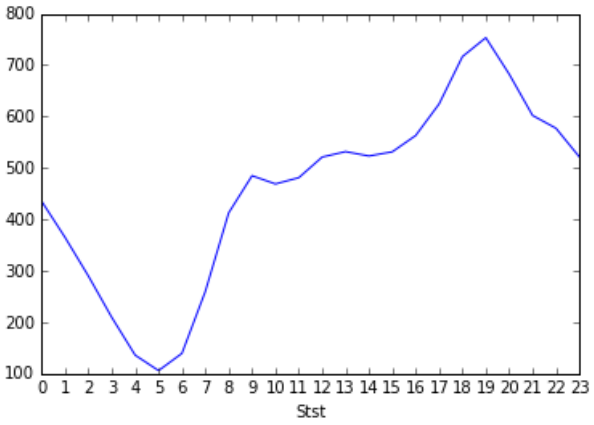


Figure 8a: Mean trip count by hour of the day

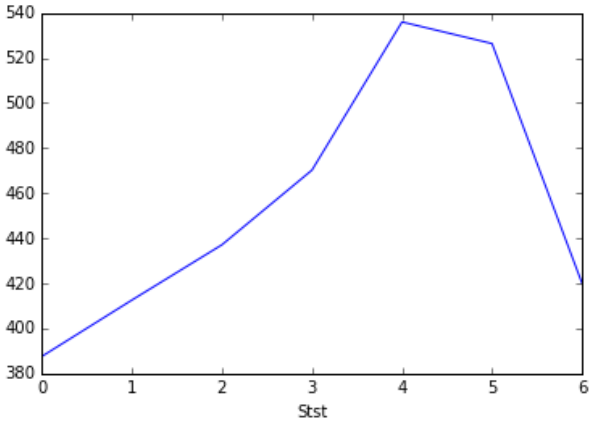


Figure 8b: Mean trip count by day of week

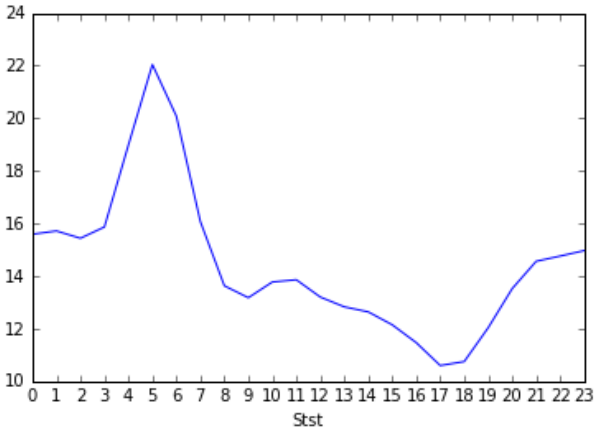


Figure 9a: Mean speeds by hour of the day

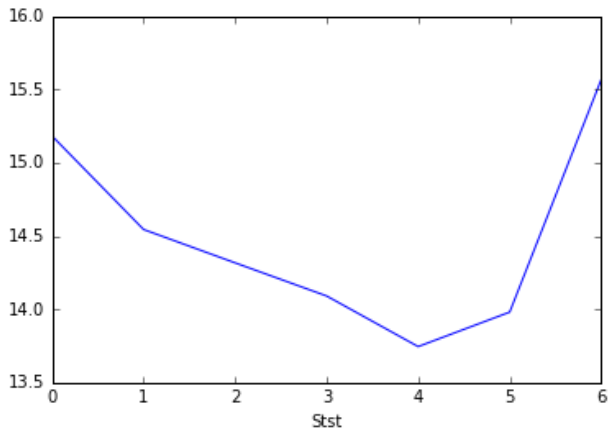


Figure 9b: Mean speeds by day of week

	Hourly		Daily		
Sr. No.	**Relationship**	**Pearson coefficient**	**Significant(p<0.05)**	**Pearson coefficient**	**Significant(p<0.05)**
1	Mean Total Fare-Mean Speed	-0.79	Yes	-0.87	Yes
2	Mean Total Fare-Mean Count	0.92	Yes	0.98	Yes
3	Mean Speed-Mean Count	-0.81	Yes	-0.86	Yes

Table 6: Pearson correlation coefficients for comparison between Mean Total Fare, Mean Speed and Mean Trip counts

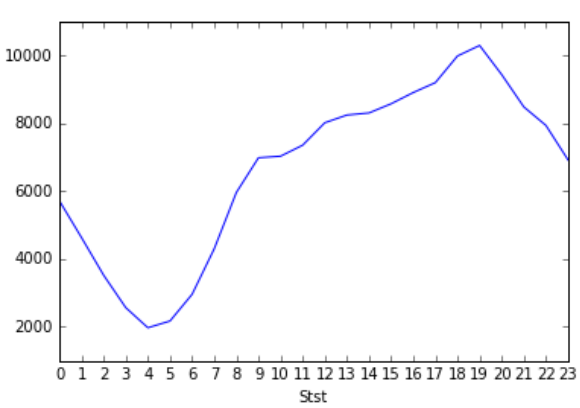


Figure 10a: Mean total fare in $ by hour of the day

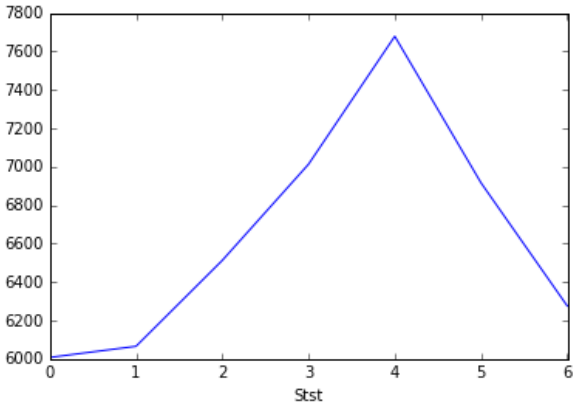


Figure 10b: Mean total fare in $ by day of week

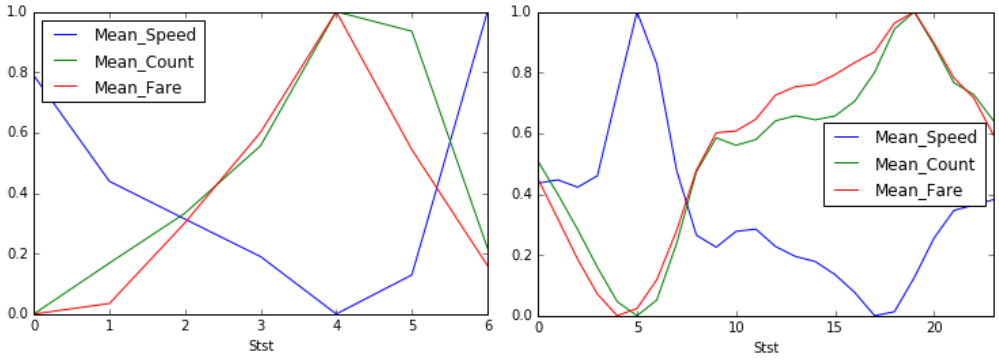
Mean trip count: As seen from the Figure 8a, the hourly mean trip count drops between 12 am (Hour 0) to 5 am when it is at the lowest. It rises steadily until it hits the peak from 5pm to 7 pm, after which it subsides until 12 am. As far as daily mean trip count is concerned, the mean number of trips is the lowest on Day 0(Monday) and highest on Day 4(Friday) as seen from Figure 8b.

Mean speeds: As seen from the Figure 9a the hourly average speeds are the fastest at 5am and the slowest from 5pm to 6 pm. As far as the daily mean speed trends go (Figure 9b), the speeds are the fastest on Sundays (Day 6) and slowest on Friday (Day 4).

From the hourly mean speeds and trip counts, we can't rule out that overall traffic volumes are lowest during early hours of the morning (4 pm to 6pm) and the highest between (5 pm and 7 pm). As far as the daily trends go, we can't rule out that the city traffic volume is the highest on Fridays and lowest on Monday resulting in slowest and fastest speeds overall respectively. This corroborates well with [28], which states that the best day to drive in Chicago is Monday and the worst day Friday. We can't rule out the maximum presence of non-Chicago vehicles and tourists/visitors on Fridays and the minimum presence on Monday.

Mean total fare: As far as the hourly mean fare goes (Figure 10a), the trends corroborate well with the mean trip counts. Lowest between 4 and 6 pm, highest between 5 pm and 7 pm. The daily fare trends (Figure 10b) also corroborate well with the trip count. Lowest on Monday, highest on Friday.

Pearson correlation coefficient: Table 6 indicates a significant strong positive correlation between Mean Total Fare and Mean Trip count and a significant strong negative correlation between Mean Total Fare and Mean Speed and Mean Speed and Mean Trip count for both hourly and daily patterns.



Figures 11: Standardized comparative plots of Mean Speed, Mean Count and Mean Total Fare by daily and hourly trends

Standardized trends (Figures 11): As we can see from the plot of standardized Mean Speed, Mean Count and Mean Total Fare both for both hourly as well as daily trends, we can see that the trends are similar for Mean Speed and Mean Count and opposite for Mean Speed and Mean Total Fare and Mean Count and Mean Total Fare. This matches well with the correlation results in Table 6.

4.4 Chicago airports:

4.4.1 Mean time to O'Hare and Midway airports: We plot the Time-Series for mean times to Community area 76(O'Hare) and community area 56: Garfield Ridge (Where Midway is located). As only a part of community area 56 is covered by Midway airport, we concentrate our attention on two pickup and drop-off co-ordinates (41.785998518, -87.750934289 and (41.796640334, -87.745282845), which are the center of the airport itself and center of the census tract which includes the airport parking lot.

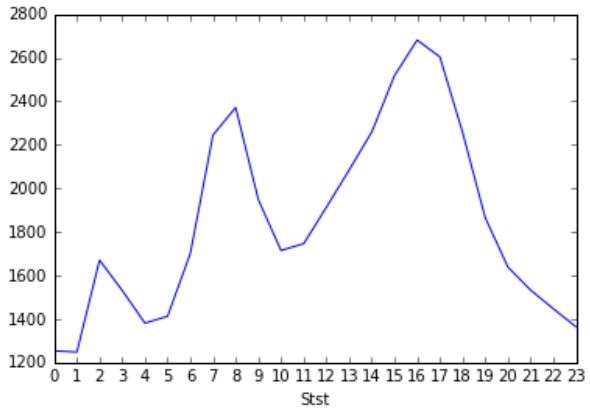


Figure 12a: Mean time to O'Hare in sec

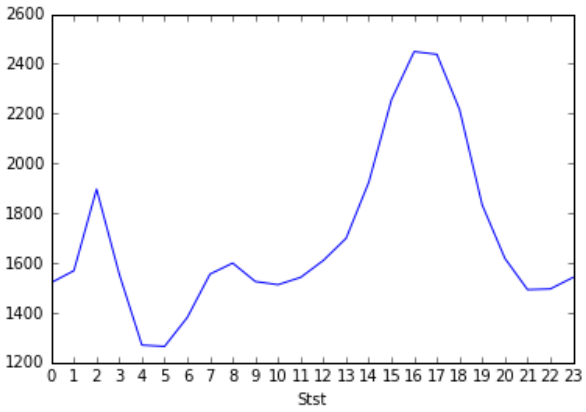


Figure 12b: Mean time to Midway in sec

Mean time to O'Hare: The overall mean time to O'Hare is the lowest from 12 am to 1 am. It rises steadily from 1 am with two local peaks in between until it hits its highest at 4 pm. Overall the worst time to travel to O'Hare is from 3 pm to 5pm and 7 am to 9 am. The best time is between 8 pm and 6 am.

Mean time to Midway: The overall mean time to Midway is the lowest from 4am to 5 am. The peak time occurs at 4 pm. Overall the worst time is from 2pm to 7 pm

4.4.2 Time from major community areas to O'Hare and Midway:

Sr. No.	Community area	Shortest(s)	Longest(s)
1	Overall	1247.93	2680.88
2	Near North Side (8)	1353.08	2769.53
3	The Loop (32)	1411.72	2780.42
4	Near West Side (28)	1280.77	2699.7
5	Near South Side (33)	1488.59	3417.27
6	Lake View (6)	1452.26	2870.59
7	Lincoln Park (7)	1333.45	2838.67

Table 7a: Mean hourly time in sec to O'Hare

Sr. No.	Community area	Shortest(s)	Longest(s)
1	Overall	1264.79	2449.19
2	Near North Side (8)	1277.70	2674.04
3	The Loop (32)	1172.90	2428.97
4	Near West Side (28)	780.0	2329.13
5	Near South Side (33)	890.0	2218.93
6	Lake View (6)	1551.43	3023.17
7	Lincoln Park (7)	1410.54	2926.69

Table 7b: Mean hourly time in sec to Midway

O'Hare route: As the route from all the community areas above uses I-90 W expressway [29], the time variations reflect the time variations and overall traffic patterns on 1-90 W.

Midway route: Similarly, as the route from all the community areas above uses I-55 expressway [29], the time variations reflect the time variations and overall traffic patterns on 1-55.

We can't rule out high traffic volumes when times are at their longest for routes to both airports above.

4.4.3. Hourly and Daily pickup and drop-off trends from/to O'Hare and Midway:

O'Hare:

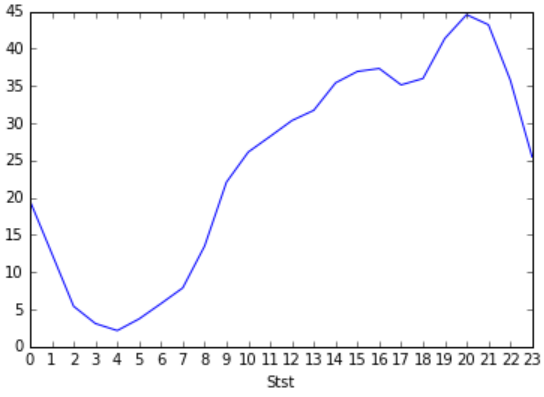


Figure 13a: Mean hourly pickups from O'Hare

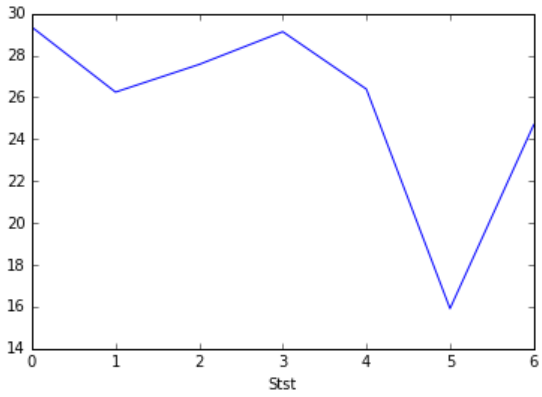


Figure 13b: Mean daily pickups from O'Hare

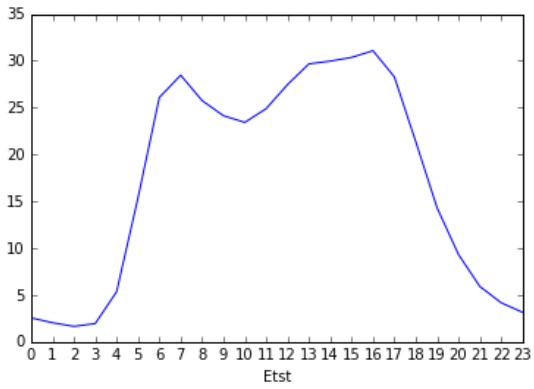


Figure 14a: Mean hourly drop-offs to O'Hare

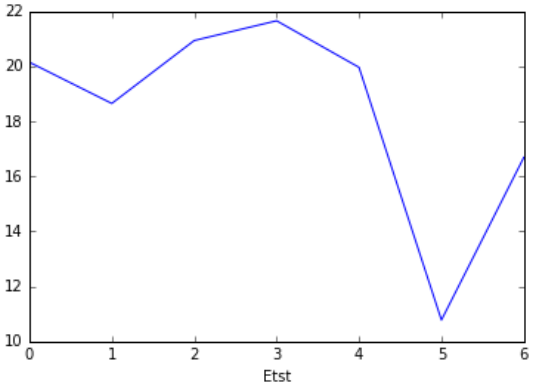


Figure 14b: Mean daily drop-offs to O'Hare

Pickup and Drop-offs patterns: As far as mean hourly pickups are considered, the least number of pickups occur between 2 am and 7 am, while the highest pickups occur between 1 pm and 10 pm. This is a fair indicator of the flight arrival patterns at O'Hare airport. Highest mean pickups occur on Monday and Wednesday and lowest on Saturdays. Highest mean hourly drop-offs occur roughly between 5 am and 7 pm. This is a fair indicator of flight departure patterns at O'Hare. Daily mean drop-offs more or the less resemble mean pickups with Thursday having the highest and Saturday having the lowest trends.

Midway:

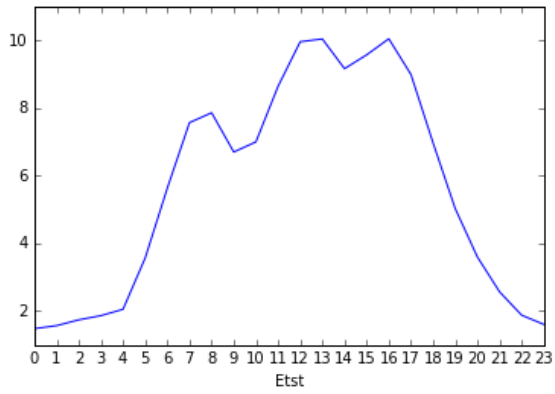


Figure 15a: Mean hourly drop-offs to Midway

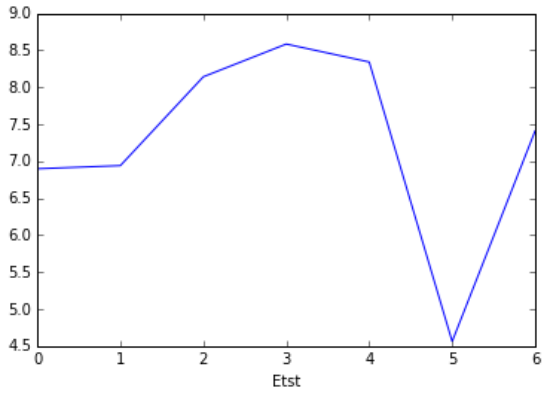


Figure 15b: Mean daily drop-offs to Midway

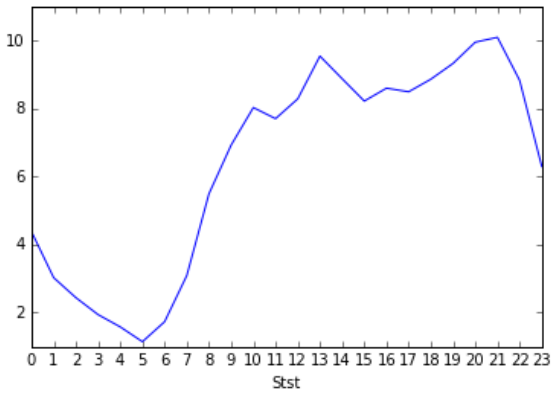


Figure 16a: Mean hourly pickups from Midway

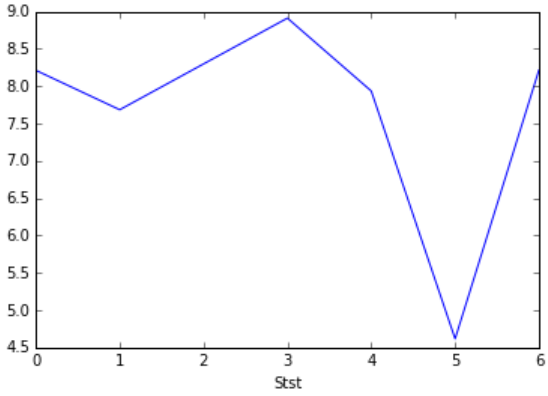


Figure 16b: Mean daily pickups from Midway

Pickup and Drop-offs patterns: As far as mean hourly pickups are considered, the least number of pickups occur between 3 am and 6 am, while the highest pickups occur between 10 am and 10 pm. This is a fair indicator of the flight arrival patterns at Midway airport. Highest mean pickups occur on Monday and lowest on Saturdays. Highest mean hourly drop-offs occur roughly between 6 am and 7 pm. This is a fair indicator of flight departure patterns at Midway. Daily mean drop-offs more or the less resemble mean pickups with Thursday having the highest and Saturday having the lowest trends.

Sr. No.	Community area	%trips
1	Near North Side	33.13
2	The Loop	21.53
3	Lake View	7.06
4	Lincoln Park	5.99
5	O'Hare	5.18
6	Near West Side	4.88
7	West Town	3.31
8	Near South Side	2.42
9	Uptown	1.78
10	Logan Square	1.68

Table 8a: Mean hourly trips from O'Hare

Sr. No.	Community area	%trips
1	Near North Side	36.47
2	The Loop	27.15
3	O'Hare	7.41
4	Lake View	5.20
5	Near West Side	5.08
6	Lincoln Park	3.88
7	Near South Side	3.29
8	West Town	2.09
9	Uptown	1.27
10	Edgewater	1.14

Table 8b: Mean hourly trips to O'Hare

O'Hare trips and community areas: The highest mean hourly pickups and drop-offs to O'Hare occur from/to the Near North Side and from/to the Loop next. However, it is interesting to know that a significant number of pickups and drop-offs are within O'Hare from/to O'Hare probably indicating International to Domestic or Domestic to International transfers**.**

Sr. No.	Community Area	%trips
1	Near North Side	40.46
2	The Loop	27.18
3	Near West Side	6.99
4	Lincoln Park	5.24
5	Near South Side	5.12
6	Lake View	4.92
7	Garfield Ridge	2.61
8	West Town	2.55
9	O'Hare	1.90
10	Uptown	0.85

Table 9a: Mean hourly trips from Midway

Sr. No.	Community Area	%trips
1	Near North Side	42.63
2	The Loop	31.92
3	Near West Side	6.23
4	Near South Side	5.61
5	Garfield Ridge	4.47
6	O'Hare	2.48
7	Lincoln Park	2.34
8	Lake View	2.32
9	West Town	0.87
10	Hyde Park	0.44

Table 9b: Mean hourly trips to Midway

Midway trips and community areas: The highest mean hourly pickups and drop-offs to O'Hare occur from/to the Near North Side and from/to the Loop next. There are considerable number of trips from/to O'Hare as well, indicating airport transfers.

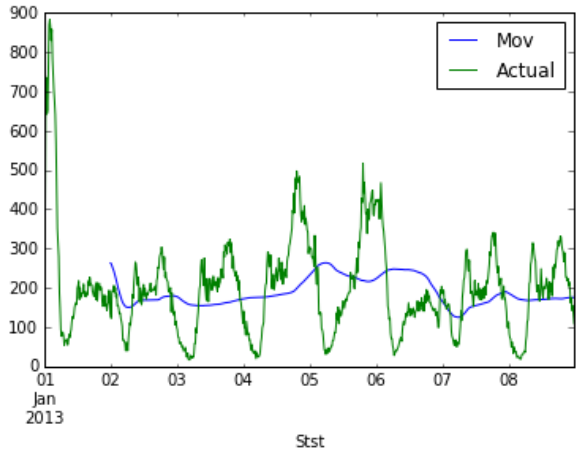
4.5 The concept of Moving Average for sections 4.6 and 4.7:

Moving average [30]: In statistics, a moving average also known as rolling average or running average is a calculation to analyze data points by creating series of averages of different subsets of the full data set. of finite impulse response filter. Variations are simple, and cumulative, or weighted forms. Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward" or, excluding the first number of the series and including the next value in the subset. A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. It can also be used for Anomaly or Event Detection [31].

Simple moving average [32]: Given a sequence $\{a_i\}_{i=1}^{N}$, an n-moving average is a new sequence $\{s_i\}_{i=1}^{N-n+1}$ defined from the a_i by taking the arithmetic mean of subsequences of n terms**,**

$$s_i = \frac{1}{n}\sum_{j=i}^{i+n-1} a_j.$$ So, the sequences S_n giving n-moving averages are

$$S_2 = \tfrac{1}{2}(a_1 + a_2, a_2 + a_3, \ldots, a_{n-1} + a_n)$$
$$S_3 = \tfrac{1}{3}(a_1 + a_2 + a_3, a_2 + a_3 + a_4, \ldots, a_{n-2} + a_{n-1} + a_n).$$



We use simple moving average for analysis of taxi trip patterns. The window size we use is 96 as there are 96, 15 minute windows in a day. This is the initial fixed subset of the number series as described above. As far as the legends go, “Mov” in the figures indicate the plot of moving averages of Taxi trip counts, “Actual” indicates the actual Taxi trip counts.

Figure 17: Example of moving average plotted against actual count trend

4.6 Patterns during Christmas, Thanksgiving and New Year's Day:

4.6.1 Patterns during Christmas:

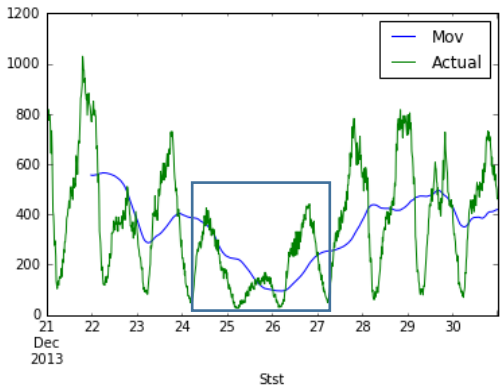


Figure 18a: Trip patterns (Dec 21 to Dec 30, 2013)

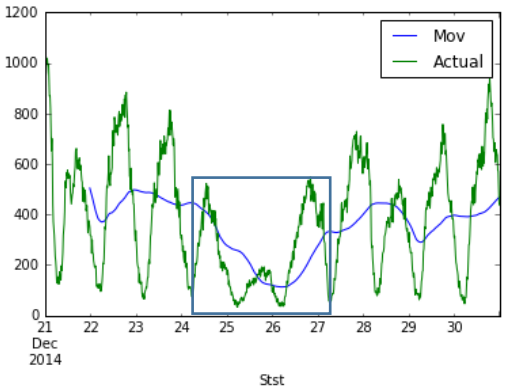


Figure 18b: Trip patterns (Dec 21 to Dec 30, 2014)

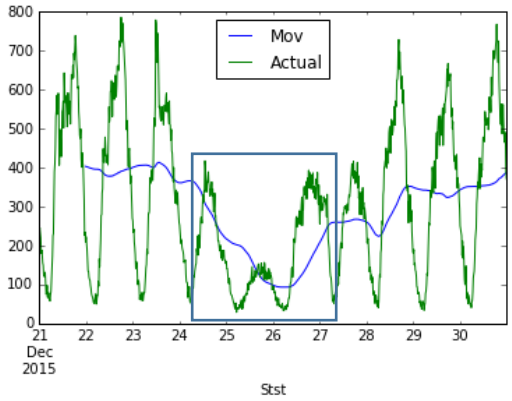


Figure 18c: Trip patterns (Dec 21 to Dec 30, 2015)

Sr. No.	Year	%Reduction between (Dec 20-23 and 24-27)
1	2013	52.12
2	2014	48.64
3	2015	48.89

Table 10: Reduction in pickups between Dec 20-23 and Dec 24-27

Sr. No.	Relationship	Pearson coefficient	Significant(p<0.05)
1	2013-2014	0.966	Yes
2	2014-2015	0.978	Yes
3	2013-2015	0.946	Yes

Table 11: Correlation between different Christmas periods (Dec 24-26)

As observed for the taxi trips for 2013, 2014 and 2015, the time series for taxi pickups from Dec 21 to Dec 30 for all the three years, we see that the taxi trips are low from Dec 24 to Dec 26, and the lowest on December 25. This indicates that not many people travel between 24 and 26. They travel the least on December 25 and prefer staying indoors on the day of Christmas. All three years recorded about 50% reduction in taxi trips between Dec 20-23 and Dec 24-27(Table 10).

Moving Average: The moving average trends for all three years also indicates a slow drop from 24th December and a gradual rise from 26th December providing a statistical basis.

Pearson correlation coefficient (Table 11): The Pearson correlation coefficient for the highlighted window (Dec 24-26) indicates a significant strong relationship between 2013, 2014 and 2015 considered pairwise. This indicates a repetitive pattern during Christmas

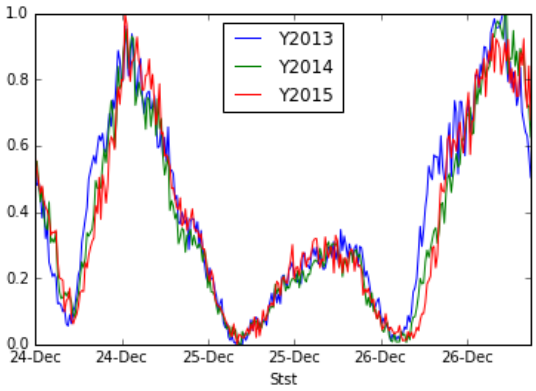


Figure 19: Comparison of taxi trip trends for years 2013, 2014 and 2015

Standardized trends (Figure 19): As we can see the trends of standardized trip counts for years 2013, 2014 and 2015 for the highlighted window (Dec 24-26) indicates strongly correlated trends, matching with the results in Table 11.

4.6.2 Patterns during Thanksgiving:

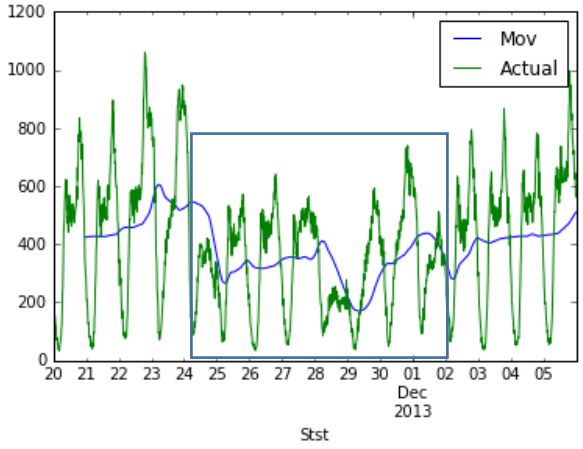


Figure 20a: Trip patterns (Nov 20 to Dec 05, 2013)

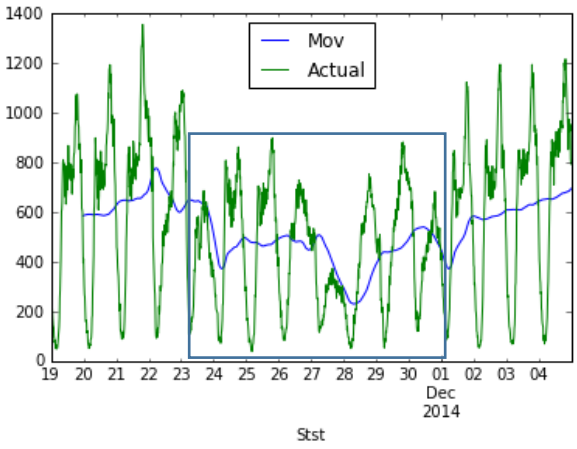


Figure 20b: Trip patterns (Nov 19 to Dec 04, 2014)

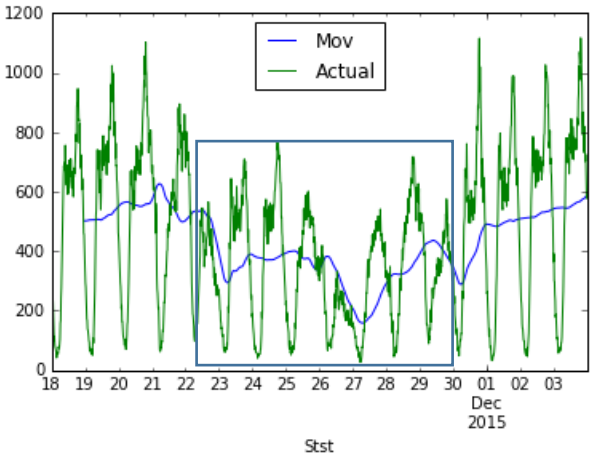


Figure 20c: Trip patterns (Nov 18 to Dec 03, 2015)

Year	Reduction between	% Reduction
2013	Reduction between Nov 14-22 and Nov 23-Dec 1	21.68
2014	Reduction between Nov 13-21 and Nov 22- 30	26.67
2015	Reduction between Nov 12-20 and Nov 21- 29	27.63

Table 12: Reduction in pickups during Thanksgiving period

Sr. No.	Relationship	Pearson coefficient	Significant(p<0.05)
1	2013-2014	0.966	Yes
2	2014-2015	0.975	Yes
3	2013-2015	0.935	Yes

Table 13: Correlation between different Thanksgiving periods

The Thanksgiving Day in 2013 was on November 28, in 2014 on November 27 and in 2015 on November 26. We can see that the taxi pickups are at their lowest on these days indicating the preference of people to stay indoors. Overall the Thanksgiving periods Nov 23 - Dec 1 for 2013, Nov 22 – Nov 30 for 2014 and Nov 21 – Nov 28 for 2015, shows 21.68%, 26.67% and 27.63% reduction from the previous 8 days (Table 12).

Moving average: The moving average trend for all years indicate that the lowest point occurs immediately at the end of the Thanksgiving Day for respective years providing statistical support.

Pearson correlation coefficient (Table 13): The Pearson correlation coefficient for the highlighted window (Blue box in figures 20a, 20b and 20c) indicates a significant strong relationship between 2013, 2014 and 2015 considered pairwise. This indicates a repetitive pattern during Thanksgiving

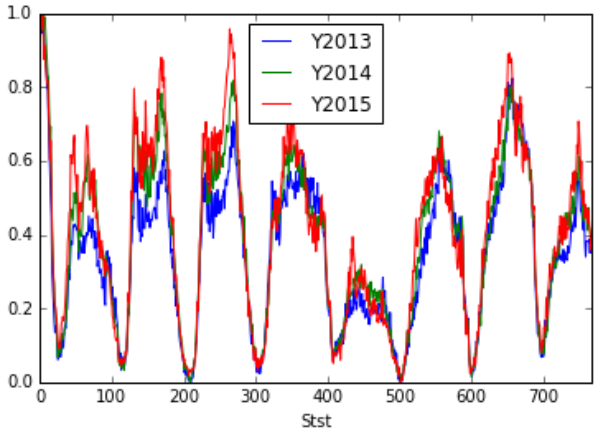


Figure 21: Comparison of taxi trip trends for years 2013, 2014 and 2015

Standardized trends (Figure 21): As we can see the trends of standardized trip counts for years 2013, 2014 and 2015 for the highlighted window (Blue box in figures 20a, 20b and 20c) indicates strongly correlated trends, matching with the results in Table 13.

4.6.3 Patterns during New Year's Day:

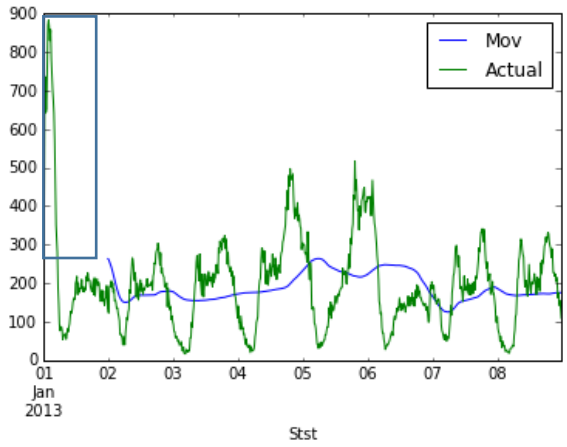


Figure 22a: Trip patterns (Jan 1 to Jan 08, 2013)

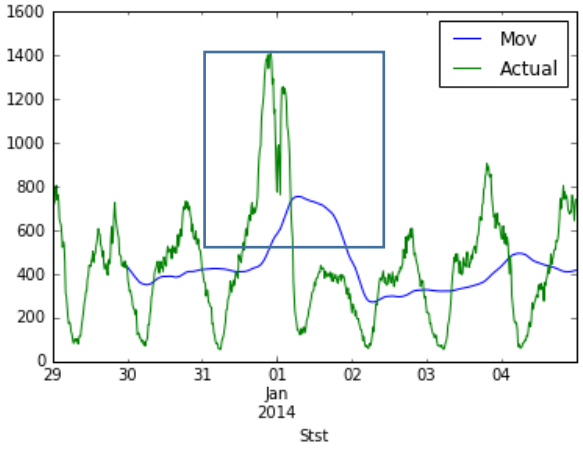


Figure 22b: Trip patterns (Dec 29,2013 to Jan 04, 2014)

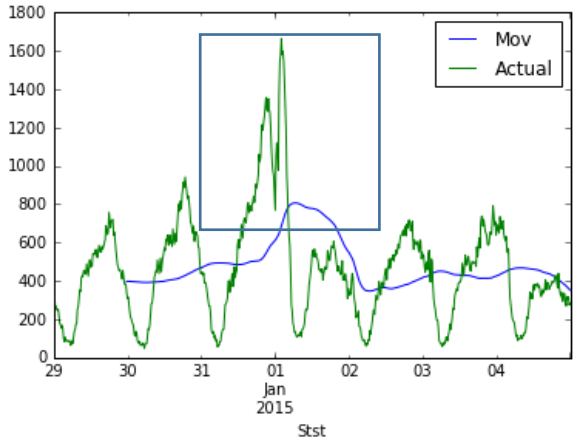


Figure 22c: Trip patterns (Dec 29,2014 to Jan 04, 2015)

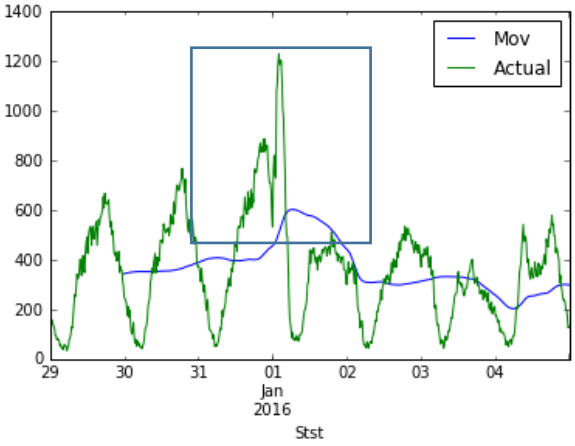
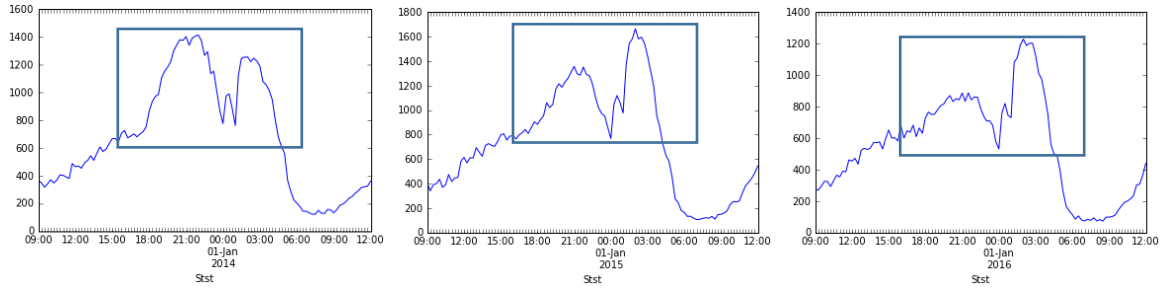


Figure 22d: Trip patterns (Dec 29,2015 to Jan 04, 2016)



Figures 23: Taxi pickup patterns between 9 am on Dec 31 and 12 pm on Jan 1 for 2014, 2015 and 2016 respectively

Sr. No.	New Year's Day	Difference	%Rise
1	2013-14	Rise between Dec 29-30 and Dec 31-Jan 1	21.90
2	2014-15	Rise between Dec 29-30 and Dec 31-Jan 1	34.08
3	2015-16	Rise between Dec 29-30 and Dec 31-Jan 1	23.05

Table 14: Statistics for rise between Dec 29-30 and Dec 31-Jan 1

Sr. No.	New Year's Eve (Dec 31 - Jan1)	Top 6 most popular drop-off community areas (descending order)
1	2013-14	8,6,32,7,28,24
2	2014-15	8,32,6,7, 28,24
3	2015-16	8,32,6,28,7,24

Table 15: Most popular drop-off community areas on New Year's days

Sr. No.	Relationship	Pearson coefficient	Significant(p<0.05)
1	2014-2015	0.966	Yes
2	2015-2016	0.984	Yes
3	2014-2016	0.934	Yes

Table 16: Correlation between different New Year periods

From figures 22 a, b, c, d, we can see that there is a surge in taxi trips between December 31 and January 1 for all 4 New Year's days 2013,2014,2015 and 2016. Upon closely observing (Figures 23), we see that there is a steady rise in the taxi pickups from about 6 pm until hits a peak from 9 pm to 10 pm. There is a temporary dip at 12.00 am until another peak between about 1 am and 4 am. This is an indicator of the pattern that people set-off for New Year's celebrations from about 6 pm to 10 pm and start returning from 1 am onwards until early hours of morning.

Moving average: For 2014, 2015, 2016, we observe the highest moving average around the morning of January 1, providing a statistical support for the observations.

Pearson correlation coefficient (Table 16): The Pearson correlation coefficient for the highlighted window (Dec 31 to January 1) indicates a significant strong relationship between 2014, 2015 and 2016 considered pairwise. This indicates a repetitive pattern during New Year's Day.

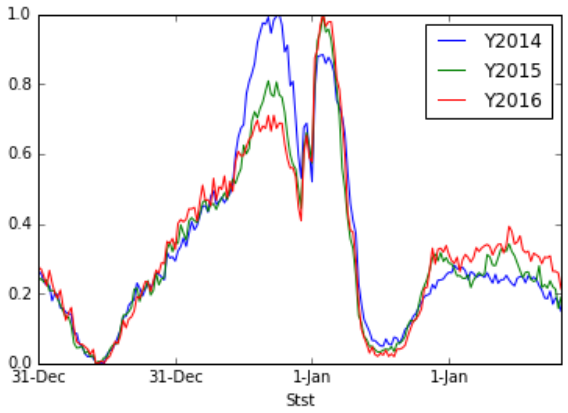


Figure 24: Comparison of taxi trip trends for years 2014, 2015 and 2016

Standardized trends (Figure 24): As we can see the trends of standardized trip counts for years 2013, 2014 and 2015 for the highlighted window (Dec 31 to January 1) indicates strongly correlated trends for major periods.

Percentage rise in taxi trips (Table 14): Between December 29-30 and December 31-January 1, the New Year's Day of 2014 saw a 21.90 % increase in number of pickups, 2015 saw 34.08 % and 2016 saw 23.05 %.

Most popular drop-off locations (Table 15): The most popular destination remained Near North side throughout, along with The Loop, Lake View, Lincoln Park, Near West Side and West Town in different orders among the top 6. This indicates the presence of various popular New Year's Day celebration places in these areas.

4.7 Patterns during the 2015 Historic Winter Storm of January 31 – February 2, 2015[33]:

The historic winter storm saw Chicago recording 19.2" snow at O'Hare [33]. The 16.2 inches recorded on February 1 was a record for any February day in Chicago [33]. Also, it was the 4th snowiest day in any month on record in Chicago [33]. Rockford recorded 11.9" of snow, and NWS Chicago and Midway airport saw Midway airport say 15.3" and 19.2" of snow respectively [33]. We analyze the impact the Snow-storm had on taxi trips and infer about the state of traffic from the patterns observed.

Impact on the number of taxi trips:

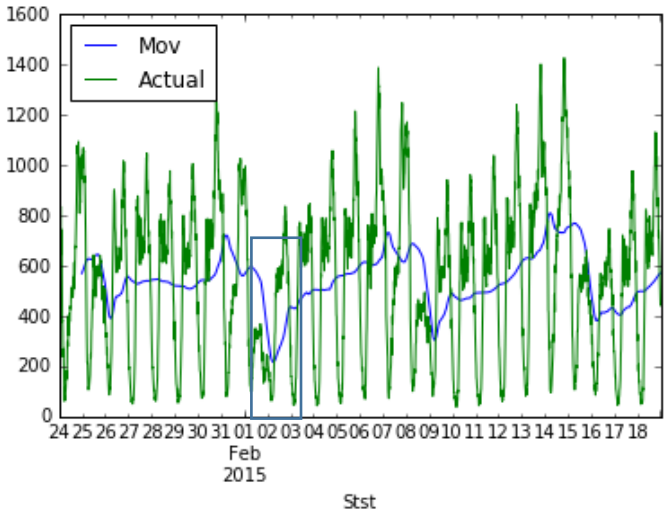


Figure 25: Taxi trip pattern in the period Jan 24, 2015 to Feb 18, 2015

Although, February 1, 2015 was a Sunday, and the peak number of Taxi trips tends to be lower compared to other days, for February 1, the peak was abnormally low compared to other Sundays, such as January 25, 2015, February 8, 2015 and February 15, 2015. Let us now see the reduction in the number of trips compared to the mean overall trips on Sundays, for total number of trips and trips by major community areas and Chicago airports.

Moving average: The lowest moving average occurs immediately at the end of February 1, 2015

Overall Reduction %	16.73

Table 17: Overall reduction in number of mean trips on Feb 1,2015 compared to overall mean trips on all Sundays.

Community area	% Reduction
Near North Side (8)	22.97
The Loop (32)	35.89
Near West Side (28)	6.21
Lake View (6)	-0.83
Lincoln Park (7)	4.73
O'Hare (76)	65.28

Table 18a: Overall reduction in mean pickups for Feb 1,2015 compared to all Sundays

Community Area	% Reduction
Near North Side (8)	29.71
The Loop (32)	36.68
Near West Side (28)	11.30
Lake View (6)	-1.08
Lincoln Park (7)	8.18
O'Hare (76)	50.14

Table 18b: Overall reduction in mean drop-offs for Feb 1,2015 compared to all Sundays

As we can see, the mean number of pickups and drop-offs show a considerable drop, especially for the Downtown areas of Chicago (Near North Side and The Loop) on the day of the blizzard. Lake View however, showed no impact of the snow storm on pickups and drop-offs. Chicago O'Hare airport recorded 65.27% drop in pickups and 50.13% drop-offs. This matches well with the fact that more than 1200 flights were cancelled at O'Hare [34], since, O'Hare sees about 2400 flights on an average every day [35]. Midway airport had a total of just 4 pickups and 40 drop-offs.

5. FUTURE RESEARCH

Future Research-An Introduction to routing using Taxisim [36]:

Taxisim is a library available under the University of Illinois/NCSA Open Source License. It is a library for Traffic Estimation, Mapping and Routing using GPS data. Overview It contains code for performing analysis of vehicle GPS data on urban road networks. Specifically, it contains traffic estimation algorithms, graph partitioning and preprocessing algorithms, efficient routing algorithms including Bidirectional Dijkstra's, Bidirectional A* and Bidirectional ArcFlags, integration with PostgreSQL[37] database systems, and a framework for performing large-scale parallel analyzes using mpi4py. The library is designed to work with maps loaded from OpenStreetMap[38] via AwesomeStitch[39]

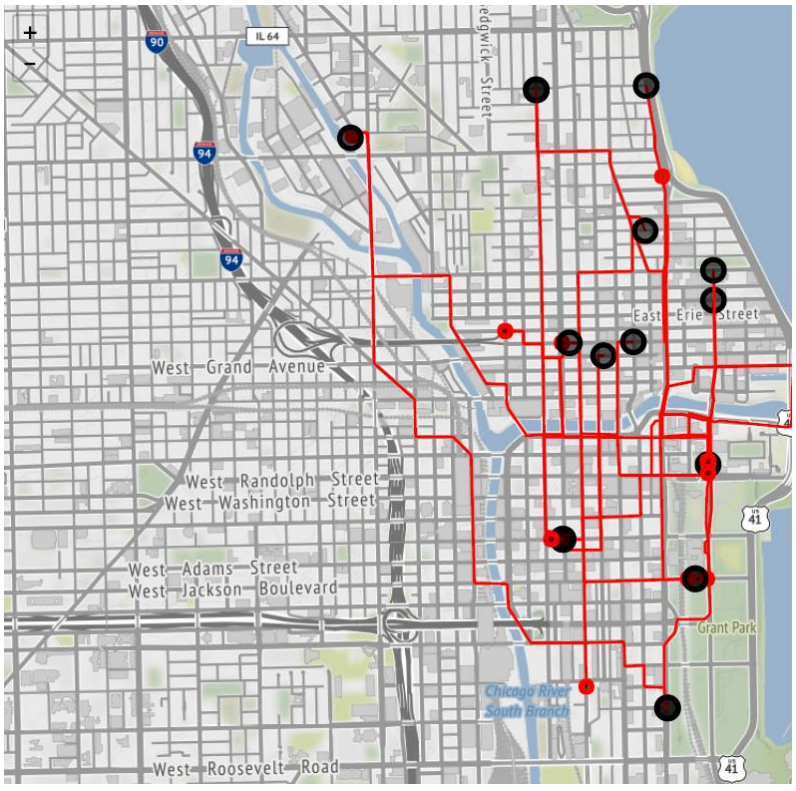


Figure 26: Routing between Community Area 8 and 32 using Folium package for Python [40]

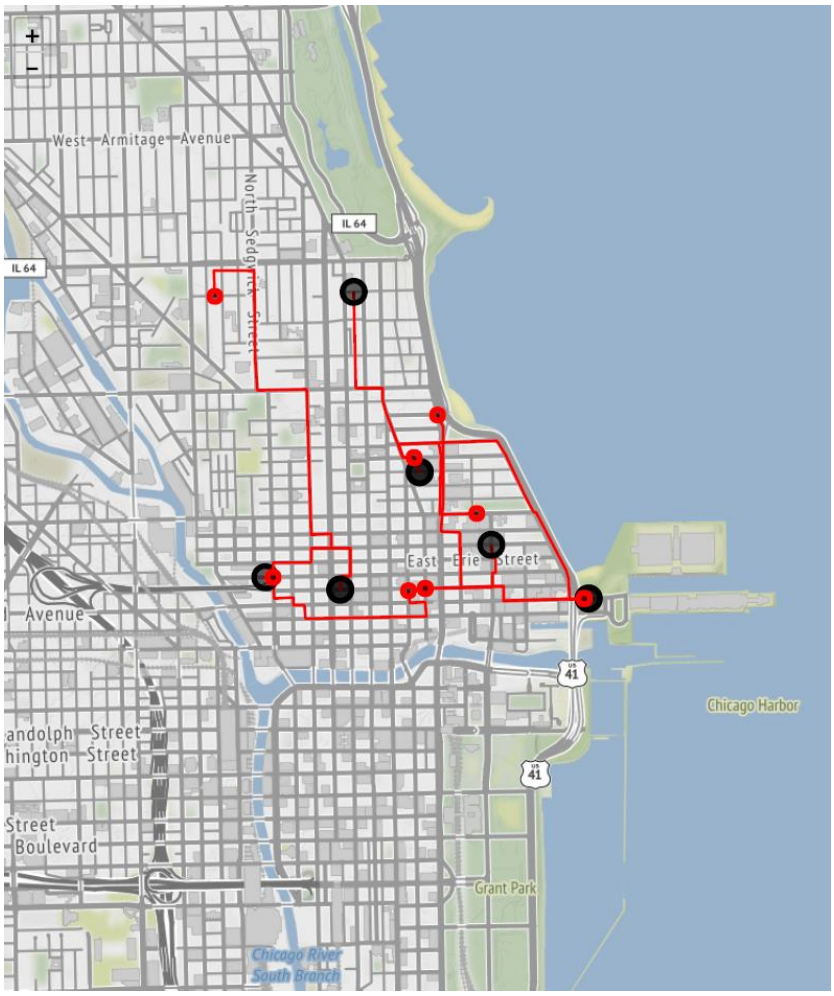


Figure 27: Routing within Community area 8 using Folium package for Python [40]

Uses: The Taxisim library [36] can be used not only for shortest path routing, but also estimating the distance and the time for the route. As the co-ordinate data is masked, the estimated routes will not be exact, but using analysis such as uncertainty quantification, the difference between the actual and the estimated time can be used for making better prediction using different modelling techniques such as regression analysis. Also, by using folium for Python [40], the routes can be visualized and can be combined with the results of this thesis to identify streets/roads such as the streets along the route to O'Hare from Community area 8(Near North Side) from the point of view identifying traffic related issues especially during times of delays.

6. CONCLUSION

From the summary statistics, the community area pairs with the highest frequency of taxi trips, the community areas with the highest percentage of pickups and drop-offs can be gauged. Also, information such as the rise/fall in demand for the top taxi companies, the mode of payment can be gauged. From hourly and weekly trends, we can infer as to which hours and days are the best and worst to travel in terms of speed and, the taxi trip trends throughout the day and the week. Also, the mean fare trends pretty much reflect the taxi trip trends in terms of rise and fall. As far as Chicago airports go, the results indicate the pickup and drop-off trends from/to Chicago airports, also the community areas with the highest inflow and outflow from/to Chicago airport and the mean time in terms of hours of the day and days of the week. This indicates the best and the worst time to travel to Chicago airport overall and from major community areas. As far as days such as Christmas, Thanksgiving and New Year's Day are concerned, we can see the taxi trends in terms of overall fall or rise compared to previous periods. The taxi trip trends during extreme weather events such as February 1, 2015 snow storm indicate the impact the weather event has on the city by looking at the drop in taxi trips overall and to/from major community areas and Chicago airports.

The trends can be useful not only for the people of Chicago, who can make better decisions as to what time of the day is the best to travel and the worst to travel. Which day of the week has slower moving traffic and which day has faster moving traffic. Also, from the trends for the mean time to Chicago airport, people have more information about the times and the days of longest and the shortest mean time to the airports from major community areas. The information can also be used by taxi companies in terms of which community areas to target for better business. Also, the hourly and daily trends in the taxi trips and the total fare collected can give information as to the peak and the slack demands during hours of the day and days of the week and days such as December 25, Thanksgiving and New Year's Day. Lastly, the information can be used by the Civic authorities to resolve traffic congestion problems in major parts of the city by observing the hourly and daily speed trends.

7. REFERENCES

[1] Chicago Taxi Data Released, November 16, 2016[Accessed on November 1, 2017],

http://digital.cityofchicago.org/index.php/chicago-taxi-data-released/

[2] Community areas in Chicago [Accessed on November 1, 2017],

https://en.wikipedia.org/wiki/Community_areas_in_Chicago

[3] Brian Patrick Donovan, University of Illinois-Urbana Champaign, 2015, "TAXIS AS PERVASIVE RESILEINCE SENSORS"

[4] Umang Patel, Department of Computer Science and Engineering University of Bridgeport, USA, 2016, "NYC Taxi Trip and Fare Data Analytics using BigData"

[5] M. Omer, A. Mostashari, and R. Nilchiani, "Assessing resilience in a regional road-based transportation network," International Journal of Industrial and Systems Engineering, vol. 13, no. 4, pp. 389-408, 2013.

[6] S. E. Chang and N. Nojima, "Measuring post-disaster transportation system performance: the 1995 Kobe earthquake in comparative perspective," Transportation Research Part A: Policy and Practice, vol. 35, no. 6, pp. 475-494, 2001.

[7] W. B. Allen, D. Liu, and S. Singer, "Accesibility measures of US metropolitan areas," Transportation Research Part B: Methodological, vol. 27, no. 6, pp. 439-449, 1993.

[8] G. Qi, X. Li, S. Li, G. Pan, Z. Wang, and D. Zhang, “Measuring social functions of city regions from large-scale taxi behaviors,” in Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011, 2011, pp.
384-388.

[9] C. Chen, D. Zhang, P. Samuel Castro, N. Li, L. Sun, and S. Li, “Realtime detection of anomalous taxi trajectories from gps traces,” in Mobile and Ubiquitous Systems: Computing, Networking, and Services, A. Puiatti and T. Gu, Eds. Springer Berlin Heidelberg, 2012, vol.
104, pp. 63-74.

[10] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, \Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips," IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2149-2158, 2013.

[11] Todd W. Schneider, “Analyzing 1.1 Billion NYC Taxi and Uber Trips, with a Vengeance”

[12] Jiawei Han and Micheline Kamber” Data Mining: Concepts and Techniques”, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000. 550 pages. ISBN 1-55860-489-8[Accessed on November 1, 2107]
http://www.cs.uiuc.edu/~hanj/bk1/1intro.ppt

[13] “Maximum likelihood estimation” [Accessed on November 14, 2107]
https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

[14] “Exponential distribution” [Accessed on November 14, 2107]

https://en.wikipedia.org/wiki/Exponential_distribution

[15] “SciPy” [Accessed on November 14, 2107]

https://docs.scipy.org/doc/scipy/reference/

[16] Jiawei Han and Micheline Kamber” Data Mining: Concepts and Techniques”, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000. 550 pages. ISBN 1-55860-489-8[Accessed on November 1, 2107]

http://www.cs.uiuc.edu/~hanj/bk1/2dw.ppt

[17] Python Data Analysis Library [Accessed on November 1, 2107]

http://pandas.pydata.org/

[18] pandas.read_csv [Accessed on November 1, 2107]

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_csv.html

[19] pandas.to_datetime [Accessed on November 1, 2107]

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.to_datetime.html

[20] pandas.DataFrame.sort_index [Accessed on November 1, 2107]

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.sort_index.html

[21] pandas.concat [Accessed on November 1, 2107]

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.concat.html

[22] pandas.DataFrame.groupby [Accessed on November 1, 2107]
https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.groupby.html

[23] pandas.DatetimeIndex.hour[Accessed on November 9, 2107]
https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DatetimeIndex.hour.html

[24] pandas.DatetimeIndex.dayofweek[Accessed on November 9, 2107]
https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DatetimeIndex.dayofweek.html

[25] "Pearson correlation coefficient" [Accessed on November 21, 2107]
https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[26] "Cross-correlation" [Accessed on November 25, 2107]
https://en.wikipedia.org/wiki/Cross-correlation

[27] "Feature scaling" [Accessed on November 23, 2107]
https://en.wikipedia.org/wiki/Feature_scaling

[28] "Best and Worst Days to Drive in U.S. Metro Areas, Cities" [Accessed on November 8, 2107]
http://www.governing.com/gov-data/transportation-infrastructure/traffic-delay-by-day-metro-areas-cities.html

[29] Courtesy Google Maps [Accessed on November 5, 2107]
https://www.google.com/maps

[30] “Moving Average” [Accessed on November 14, 2107]
https://en.wikipedia.org/wiki/Moving_average

[31] https://datascience.stackexchange.com/questions/6547/open-source-anomaly-detection-in-python [Accessed on November 14, 2107]

[32] Moving Average [Accessed on November 14, 2107]
http://mathworld.wolfram.com/MovingAverage.html

[33] “Historic Winter Storm of January 31-February 2, 2015” [Accessed on November 8, 2107]
https://www.weather.gov/lot/2015_Feb01_Snow

[34] “Midwest snow storm grounds hundreds of Chicago flights” [Accessed on November 8, 2017]
https://www.reuters.com/article/us-usa-weather/midwest-snow-storm-grounds-hundreds-of-chicago-flights-idUSKBN140127

[35] CHICAGO O'HARE AIRPORT (ORD) [Accessed on November 8, 2017]
https://www.airport-ohare.com/

[36] Lab-Work/taxisim [Accessed on November 9, 2017]
https://github.com/Lab-Work/taxisim

[37] PostgreSQL [Accessed on November 9, 2017]
https://www.postgresql.org/

[38] OpenStreetMaps [Accessed on November 9, 2017]

https://wiki.openstreetmap.org/wiki/Main_Page

[39] Lab-Work/AwesomeStitch [Accessed on November 9, 2017]

https://github.com/Lab-Work/AwesomeStitch

[40] folium 0.5.0 [Accessed on November 9, 2017]

https://pypi.python.org/pypi/folium